NMR analysis of the trans-activation response (TAR) RNA element of equine infectious anemia virus

David W. Hoffman* and Stephen W. White

Department of Chemistry and Biochemistry, University of Texas at Austin, Austin, TX 78712, USA and Department of Microbiology, Duke University Medical Center, Durham, NC 27710, USA

Received July 12, 1995; Revised and Accepted September 11, 1995

ABSTRACT

Transcription of lentiviral DNA in the host cell is regulated by an interaction between the viral TAR RNA stem–loop and the viral Tat protein. Here we present a model of the three-dimensional structure of the TAR RNA stem–loop of the equine infectious anemia virus (EIAV), derived from two- and three-dimensional NMR data. This 25 nucleotide RNA consists of an A-form helical stem capped by two U-G base pairs and a four-nucleotide loop. Two loop cytidines are stacked into the loop interior and likely form a non-Watson–Crick C-C base-pair. The two nucleotides at the top of the loop, U13 and G14, appear to be excluded from the interior of the loop and solvent exposed. It is significant that now for the EIAV TAR-Tat system, three-dimensional structures are now known for both the RNA and protein components.

INTRODUCTION

Many of the lentivirinae subfamily of retroviruses utilize a novel protein–RNA interaction that promotes the transcription of viral RNA during infection. The system was first discovered in HIV-1 (1–3) but was subsequently shown to be present in related viruses that compromise the immune systems of animal groups such as cattle (BIV), cats (FIV), monkeys (SIV) and horses (EIAV) (4–6). The system consists of two basic components, a virally encoded trans-activator protein or Tat that specifically binds a region of viral RNA called the Tat-responsive element or TAR (7,8). Its operation in HIV-1 has attracted great interest because it is essential for viral survival and is a potential target for drugs to treat AIDS. The HIV-1 Tat protein contains 86 amino acids and can be divided into five regions of which the most important are the core and basic regions (4,9–11). The TAR element is a 60-nucleotide stem–loop structure located at the 5’ end of viral RNA (12). Tat binds to TAR via its basic region (13–15) and the interaction results in increased levels of viral mRNA (1–3). The exact mechanism by which this interaction stimulates viral transcription is unclear; suggestions include increased efficiency and/or processivity (16–20), anti-termination (21) and modification of the transcription complex (5). Other evidence suggests that the effect is elicited by a host factor that binds to Tat and is thereby brought to the transcription complex (22–24).

There are practical advantages to studying the Tat-TAR system in a non-human system. Equine infectious anemia virus, like HIV, is a member of the lentivirinae subfamily, and is the causative agent of EIAV in horses. Equine infectious anemia virus and HIV-1 are similar in the organization of their genomes, their inclination to attack the immune system of the host, and their mechanisms for regulating gene expression. Unlike HIV-1, however, EIAV is usually contained by the immune system of the host and is usually not fatal to the infected animal. These factors combine to make EIAV an important model system for understanding and eventually controlling retroviral infection. The Tat-TAR system of EIAV has also been well characterized, and although there are differences from the HIV-1 system, the similarities suggest that the underlying mechanism is the same. The 75 amino acid Tat protein from EIAV lacks a cysteine-rich region present in other Tat proteins but does contain the important core and basic regions (4,25). In a series of experiments involving EIAV and HIV-1 Tat proteins, and chimeras between the two, Derse and coworkers have demonstrated their functional equivalence, mapped their RNA-binding and activation regions and shown that they probably bind the same cellular component (5,24–26). The solution structure of EIAV Tat has also been determined which confirms many of the predictions from the amino acid sequence (27). Regarding the EIAV TAR element, it was predicted to comprise a 25-nucleotide stem–loop (Fig. 1) and this structure is supported by site-specific mutagenesis experiments (28), as well as our own previous low-resolution NMR results (29).

Since the EIAV TAR RNA is much smaller than its HIV-1 counterpart, it is considerably more amenable to structure analysis by NMR methods. We recently initiated such a study and determined the general structural features of the EIAV TAR RNA stem–loop from two-dimensional NMR data and nuclease cleavage experiments (29). We have now significantly extended our earlier studies, and report here a three-dimensional structural model of the RNA molecule derived from two- and three-dimensional NMR data. The results of our analysis, when combined with previously published mutational results, suggest the RNA nucleotides that are most likely to be important in the EIAV Tat-TAR system. This work represents a continuation of our* To whom correspondence should be addressed
dimensional NMR spectra were acquired in the phase-sensitive quantum coherence (HMQC) spectrum. All two- and three-mode by the method of time proportional phase incrementation determined at natural abundance from a heteronuclear multiple mixing times of 50 and 80 ms. Carbon chemical shifts were acquired both with and without phosphorous decoupling during times of 80, 120, 160, 200 and 280 ms. 2QF-COSY spectra were effect (NOE) spectra in D2O solvent were acquired with mixing.

NMR experiments

EIAP TAR RNA. NMR data were collected at 500 or 600 MHz using either Varian Unity (500 or 600 MHz) or 500 MHz Bruker AMX spectrometers. Spectra in D2O solvent were obtained using samples with 5-12 mg of RNA dissolved in 550 µl of 15 mM Na/K phosphate buffer at pH 6.0 and 300 nM template DNA and typically yielded 1 mg of purified RNA per 20 ml transcription volume. The RNA was purified using polyacrylamide gel electrophoresis under denaturing conditions (8 M urea), electroeluted, ethanol precipitated and finally passed through a Sephadex G25 gel filtration column in 2 mM phosphate buffer and lyophilized.

RNA structure determination

The structure determination of the EIAV TAR RNA was carried out using the hybrid distance geometry-simulated annealing and energy minimization protocols within the X-PLOR version 3.1 program suite (37). The target function that was minimized during the simulated annealing process contained the following: (i) square-well potentials for interproton distance constraints and torsion angle constraints; (ii) hydrogen bond distances for the two U-G base pairs were constrained to be <2.4 Å; (iii) quadratic harmonic potential terms were used for covalent geometry (including bonds, angles, planes and chirality). The simulated annealing was carried out using the standard force-field parameter set (parallel.dna). No electrostatic terms where included in the target function. The coordinates of nucleotides 10–17 were varied during the simulated annealing process, while the structure of nucleotides 3–9 and 18–24 were restricted to A-form helical RNA (38). Nucleotides 1, 2 and 25 at the 5’ and 3’ ends of the RNA were not included in the structure analysis. A total of 92 relatively ‘loose’ distance constraints were derived from observed NOEs and used in defining our model of the eight nucleotide loop. Fifty-two of these constraints were derived from an NOE spectrum acquired for a 2 mM sample at 27°C, in 10 mM phosphate buffer at pH 6.0, with a relatively short mixing time of 80 ms to minimize the effects of spin diffusion on the NOE intensities. Depending on its intensity, each NOE was placed into one of four categories ranging from very strong to weak. Each intensity category was then assigned to an interproton distance range of <2.8 Å, <3.5 Å, 2.4–4.2 Å and 3–4.5 Å, respectively. Thirty-four additional NOEs were detected in NOE spectra with mixing times of 120, 160 or 200 ms, and these were assigned to distance ranges of 3.5–6 Å. Six additional distance constraints were derived from peaks that were observed in the three-dimensional NOEY-TOCSY spectrum, but were obscured by overlap in all of the two-dimensional data sets. These were assumed to correspond to an interproton separation of <4.5 Å for the NOE transfer. No intra-ribose or intra-base NOE derived distance constraints were used. It was possible to internally verify these NOE/distance assignments using NOE data from the A-form...
stem where the interproton distances are known. Torsion angle
restraints were used to restrict the riboses of U10, G16 and G17
to the C3' endo conformation, based on their small H1'-H2'
intraribose coupling constants. The torsion angles of the other
loop riboses were not restricted. Hydrogen bonding restraints for
the two U-G base pairs were included, in lieu of using NOEs
involving the exchangeable protons. The hydrogen bond dis-
tances were restricted to be <2.4 Å and the bases in each U-G pair
was restricted to be within 10° of co-planar. To ensure that a wide
range of conformational space was being sampled during the
simulated annealing process, a variety of starting structures were
used in which the initial positions of the loop nucleotides were far
from their final positions. A set of 10 structures were calculated
for each of five different starting structures, for a total of 50
structures calculated via the simulated annealing process. Struc-
tures with no violations of our NMR-derived distance and angle
constraints and minimum values of the X-PLOR target function
were found to contain similar features. For the 20 lowest energy
structures, the average root mean square difference is 1.6 Å for
loop nucleotides 10–17. The average root mean square difference
is as high as 4.5 Å for atoms on the bases of U13 and G14.

RESULTS

NMR analysis of the EIAV TAR RNA

A detailed investigation of a macromolecular structure using
NMR methods requires that as many resonances as possible be
unambiguously assigned to the protons from which they orig-
ninate. This assignment problem is particularly difficult in the case
of RNA due to the extensive resonance overlap, particularly
among the ribose protons. Using two-dimensional NMR methods
that are now considered rather standard, we were previously
successful in obtaining partial resonance assignments for the
EIAV TAR RNA stem–loop (29). These assignments were
sufficient to derive an essentially two-dimensional model de-
scribing the base pairing and sugar conformations of the
stem–loop. In the work we present here, we use additional NMR
approaches to largely overcome the resonance overlap problem
and obtain a significantly increased number of resonance
assignments (Table 1). These assignments have allowed us to
identify additional interproton distance constraints (summarized in Table 2) that were then used to derive a full three-dimensional
model of the EIAV TAR RNA structure.

The NMR technique that proved most useful for obtaining
additional resonance assignments was homonuclear three-dimen-
sional NOESY-TOCSY. This technique has previously been used
as a tool in studying proteins (34) and DNA (35) and has only
recently been applied to RNA (36). In the three-dimensional
NOESY-TOCSY spectrum, peaks in the two-dimensional NOE
spectrum are resolved into a third dimension by the chemical shift
of a proton which is scalar (through-bond) coupled to one of the
protons in the pair giving rise to the NOE (Fig. 2). Figure 3
illustrates the usefulness of the three-dimensional NOESY-
TOCSY spectrum in identifying and resolving additional NOEs
that were obscured by overlap in the two-dimensional spectra. As
can be seen in the spectra in Figure 3, the three-dimensional
NOESY–TOCSY spectrum exhibits quite reasonable resolution
and excellent signal-to-noise.

The assignments derived from the three-dimensional NOESY–
TOCSY spectrum were used to identify corresponding NOE
peaks in the two-dimensional NMR data. The intensities of the
latter were then evaluated and translated into interproton
distances to be used as structural constraints. This procedure was
adopted because the peak intensities in the three-dimensional
spectrum do not provide distance bounds that are as precise as
those obtained from higher resolution two-dimensional data. The
reason for this limitation is that each peak in the three-dimen-
sional spectrum is a result of two transfers (one through bond and
one through space) of unequal and variable efficiency. In several
instances, peaks assigned using the three-dimensional NOESY–
TOCSY spectrum could not be unambiguously identified with
the corresponding peaks in the two-dimensional NOE spectrum due
to overlap with other resonances. In these cases, we used the
three-dimensional data directly to obtain interproton distance
estimates. Although less precise, they do provide additional
‘upper limit’ distance constraints that are valuable for further
defining the structure.

We also employed another approach in our efforts to obtain the
maximum number of interproton distance constraints. A series of
two-dimensional NOESY, TOCSY and 2QF-COSY spectra were
collected at five temperatures between 2 and 30°C. The chemical
shifts of many of the protons were found to exhibit small but
significant temperature dependence in the order of 0.02
p.p.m./10°C. This temperature dependence was sufficient in
many cases to separate peaks that overlap at a particular
temperature. Given the small size of the shifts, they are unlikely
to represent a significant structural change in the RNA. They are
more likely to be a consequence of the hypersensitivity of
chemical shifts to very small conformational changes. The
hypothesis that there is little structural change in the EIAV TAR
RNA over the investigated temperature range is supported by the
relative intensities of the NOEs, which show minimal variation
with temperature.

Description of the RNA structure

RNA structure determination by NMR methods can be particular-
ly difficult due to the relatively low density of NMR-observable
protons, especially in comparison to proteins. Even when one
is able to largely solve the overlap problem as we have
through the use of homonuclear three-dimensional NMR
methods, we still must contend with the difficulty of having to
define the positions of the loop nucleotides using a relatively
modest density of distance constraints. As the most extreme
example, the location of the purine ring of G14 must be
determined using NOEs involving only one proton (the H8), since
the imino and amino protons of G14 are exchanging with the
solvent too quickly to be observed and are apparently not
hydrogen bonded. The density of NMR-derived distance con-
straints ultimately limits the precision of our structural models.
However, our available constraints clearly contain a substantial
amount of structural information regarding the relative positions
of the loop nucleotides, and suggest the interactions that stabilize
the loop. When combined with previously reported mutational
results, our structural model suggests which nucleotides are
available for interaction with the EIAV Tat protein or as yet
unidentified factors.

The structure of the EIAV TAR RNA stem–loop can be divided
into three distinct regions: a seven base pair stem, an eight
nucleotide loop and the nucleotides at the 5' and 3' ends of
the molecule (Fig. 1). These will be described in turn.
This analysis was facilitated by the relatively wide distribution to interproton distance bounds and torsion angles, respectively. Thus, the relevant NOEs and coupling constants derived and separation of the proton chemical shifts in the loop as were evaluated from the NMR spectra, and these were converted therefore restricted to that of a standard A-form helix. The geometry of the stem nucleotides was based on the constraints from our NMR model the stem ab initio. The NMR data from this region are entirely consistent with an A-form Watson–Crick base paired helical structure. For example, the U-G base pairs.

The stem. The RNA stem comprises nucleotides 3–9 and 18–24. The NMR data from this region are entirely consistent with an A-form Watson–Crick base paired helical structure. For example, the riboses are in the C3’ endo conformation since they exhibit the characteristic H1′-H2′ coupling constants (Table 1). Also, the relative intensities of all the assigned NOEs are consistent with the interproton distances of the A-form helix. Overall, the chemical shifts for the protons in the stem show a rather narrow range of structures that were obtained. The loop can be subdivided into three structural regions: the two U-G pairs at the base of the loop, a C-C pseudo base pair and the two nucleotides at the turn.

The RNA loop. The loop region comprises nucleotides 10–17. Unlike the stem region, the NMR data from the loop do not correspond to a regular structure and its conformation was derived ab initio. Thus, the relevant NOEs and coupling constants were evaluated from the NMR spectra, and these were converted to interproton distance bounds and torsion angles, respectively. This analysis was facilitated by the relatively wide distribution and separation of the proton chemical shifts in the loop as compared to the stem. Examples of the RNA loop structures are shown in Figures 4 and 5. Figure 4 shows a set of five of the 50 total energy minimized structures that are representative of the range of structures that were obtained. The loop can be subdivided into three structural regions: the two U-G pairs at the base of the loop, a C-C pseudo base pair and the two nucleotides at the turn.

### Table 1. Assignments of chemical shifts (p.p.m.) and intra-ribose coupling constants (Hz) for the EIAV TAR RNA in 10 mM phosphate buffer, pH 6

<table>
<thead>
<tr>
<th>H8/H6</th>
<th>C8/C6</th>
<th>H5/H2</th>
<th>C5/C2</th>
<th>H1'</th>
<th>H2'</th>
<th>H3'</th>
<th>H4'</th>
<th>H5'</th>
<th>J(1'-'2')</th>
<th>imino/amino</th>
</tr>
</thead>
<tbody>
<tr>
<td>G1</td>
<td>7.95</td>
<td>138.2</td>
<td>na</td>
<td>na</td>
<td>5.63</td>
<td>4.71</td>
<td>4.43</td>
<td>4.74</td>
<td>4.22</td>
<td>4.15</td>
</tr>
<tr>
<td>C2</td>
<td>7.87</td>
<td>143.0</td>
<td>5.96</td>
<td>97.2</td>
<td>5.87</td>
<td>4.41</td>
<td>4.58</td>
<td>4.52</td>
<td>4.19</td>
<td>4.21</td>
</tr>
<tr>
<td>A3</td>
<td>8.30</td>
<td>139.7</td>
<td>7.94</td>
<td>153.3</td>
<td>5.90</td>
<td>4.75</td>
<td>4.15</td>
<td>4.55</td>
<td>4.22</td>
<td></td>
</tr>
<tr>
<td>C4</td>
<td>7.51</td>
<td>140.0</td>
<td>5.27</td>
<td>95.9</td>
<td>5.48</td>
<td>4.32</td>
<td>4.49</td>
<td>4.17</td>
<td>&lt;3</td>
<td>8.44, 7.01</td>
</tr>
<tr>
<td>U5</td>
<td>7.86</td>
<td>141.0</td>
<td>5.35</td>
<td>102.0</td>
<td>5.56</td>
<td>4.49</td>
<td>&lt;3</td>
<td>14.10</td>
<td>&lt;3</td>
<td>8.27, 6.97</td>
</tr>
<tr>
<td>C6</td>
<td>7.79</td>
<td>140.2</td>
<td>5.65</td>
<td>96.4</td>
<td>5.49</td>
<td>4.46</td>
<td>4.53</td>
<td>&lt;3</td>
<td>8.00, 6.75</td>
<td></td>
</tr>
<tr>
<td>A7</td>
<td>7.96</td>
<td>138.2</td>
<td>6.86</td>
<td>151.5</td>
<td>5.84</td>
<td>4.43</td>
<td>4.64</td>
<td>4.14</td>
<td>&lt;3</td>
<td></td>
</tr>
<tr>
<td>G8</td>
<td>7.11</td>
<td>134.4</td>
<td>na</td>
<td>5.53</td>
<td>4.53</td>
<td>4.43</td>
<td>4.05</td>
<td>&lt;3</td>
<td>12.50</td>
<td></td>
</tr>
<tr>
<td>A9</td>
<td>7.57</td>
<td>137.9</td>
<td>7.83</td>
<td>153.1</td>
<td>5.91</td>
<td>4.62</td>
<td>4.41</td>
<td>4.47</td>
<td>4.07</td>
<td>&lt;3</td>
</tr>
<tr>
<td>U10</td>
<td>7.47</td>
<td>139.0</td>
<td>5.26</td>
<td>102.8</td>
<td>5.33</td>
<td>4.11</td>
<td>4.38</td>
<td>4.46</td>
<td>4.03</td>
<td>3</td>
</tr>
<tr>
<td>U11</td>
<td>7.88</td>
<td>141.5</td>
<td>5.60</td>
<td>103.4</td>
<td>5.77</td>
<td>4.28</td>
<td>4.52</td>
<td>4.36</td>
<td>4.05</td>
<td>7</td>
</tr>
<tr>
<td>C12</td>
<td>7.76</td>
<td>142.7</td>
<td>5.92</td>
<td>96.8</td>
<td>5.62</td>
<td>4.49</td>
<td>4.40</td>
<td>4.31</td>
<td>4.07</td>
<td></td>
</tr>
<tr>
<td>U13</td>
<td>7.73</td>
<td>142.0</td>
<td>5.60</td>
<td>103.4</td>
<td>5.53</td>
<td>4.34</td>
<td>4.37</td>
<td>4.11</td>
<td>3.92</td>
<td>4.08</td>
</tr>
<tr>
<td>C14</td>
<td>7.80</td>
<td>137.3</td>
<td>na</td>
<td>5.60</td>
<td>4.49</td>
<td>4.56</td>
<td>4.31</td>
<td>3.94</td>
<td>4.13</td>
<td></td>
</tr>
<tr>
<td>C15</td>
<td>7.77</td>
<td>142.0</td>
<td>5.60</td>
<td>96.1</td>
<td>5.83</td>
<td>4.46</td>
<td>4.58</td>
<td>4.39</td>
<td>4.09</td>
<td>4.22</td>
</tr>
<tr>
<td>G16</td>
<td>8.00</td>
<td>137.8</td>
<td>na</td>
<td>5.49</td>
<td>4.74</td>
<td>4.56</td>
<td>4.46</td>
<td>4.08</td>
<td>&lt;3</td>
<td>10.82</td>
</tr>
<tr>
<td>G17</td>
<td>7.42</td>
<td>136.2</td>
<td>na</td>
<td>5.65</td>
<td>4.63</td>
<td>4.33</td>
<td>4.14</td>
<td>&lt;3</td>
<td>11.75</td>
<td></td>
</tr>
<tr>
<td>U18</td>
<td>7.75</td>
<td>140.2</td>
<td>5.29</td>
<td>101.8</td>
<td>5.54</td>
<td>4.33</td>
<td>4.48</td>
<td>&lt;3</td>
<td>14.42</td>
<td></td>
</tr>
<tr>
<td>C19</td>
<td>7.82</td>
<td>140.3</td>
<td>5.66</td>
<td>96.4</td>
<td>5.54</td>
<td>4.37</td>
<td>4.43</td>
<td>&lt;3</td>
<td>8.33, 7.11</td>
<td></td>
</tr>
<tr>
<td>U20</td>
<td>7.76</td>
<td>142.7</td>
<td>5.38</td>
<td>102.4</td>
<td>5.47</td>
<td>4.58</td>
<td>4.44</td>
<td>&lt;3</td>
<td>12.27</td>
<td></td>
</tr>
<tr>
<td>G21</td>
<td>7.68</td>
<td>135.1</td>
<td>na</td>
<td>5.59</td>
<td>4.57</td>
<td>4.48</td>
<td>4.15</td>
<td>&lt;3</td>
<td>11.65</td>
<td></td>
</tr>
<tr>
<td>A22</td>
<td>7.72</td>
<td>138.0</td>
<td>7.36</td>
<td>152.4</td>
<td>5.85</td>
<td>4.59</td>
<td>4.61</td>
<td>4.47</td>
<td>4.12</td>
<td>&lt;3</td>
</tr>
<tr>
<td>G23</td>
<td>7.08</td>
<td>134.3</td>
<td>na</td>
<td>5.56</td>
<td>4.36</td>
<td>4.65</td>
<td>4.00</td>
<td>&lt;3</td>
<td>13.55</td>
<td></td>
</tr>
<tr>
<td>U24</td>
<td>7.52</td>
<td>140.0</td>
<td>5.02</td>
<td>101.9</td>
<td>5.66</td>
<td>4.33</td>
<td>4.37</td>
<td>4.04</td>
<td>&lt;3</td>
<td></td>
</tr>
<tr>
<td>C25</td>
<td>7.81</td>
<td>142.1</td>
<td>5.80</td>
<td>96.9</td>
<td>5.77</td>
<td>4.03</td>
<td>4.18</td>
<td>&lt;3</td>
<td></td>
<td>9</td>
</tr>
</tbody>
</table>

Data for non-exchangeable protons are reported at 30°C, with chemical shifts relative to the residual HDO resonance at 4.7 p.p.m.. Carbon chemical shifts are relative to external TMS. Intra-ribose coupling constants J(H1′-H2′) that were undetectable are reported as <3 Hz, other couplings have an uncertainty of 1 Hz. Amino and imino protons are reported at 2°C where their exchange rate with the solvent is reduced.
would have been in a completely regular A-form helix with only Watson–Crick type base pairs.

H1′ to H2′ scalar couplings indicate that the riboses of U10, G16 and G17 are in the C3′ endo conformation, but the ribose of U11 is primarily C2′ endo. The C2′ endo conformation usually occurs in extended regions because it increases the sequential interphosphate distance (40–42). In the EIAV TAR RNA stem loop, this extended conformation most likely facilitates the transition from the stem region of the RNA into the tight turn at the top of the loop.

The U-G base pairs in the EIAV TAR RNA are similar to those reported in other RNA structures. The RNA stem–loop investigated in solution by NMR methods by Puglisi and coworkers (42) also contains a U-G pair at the top of an A-form helical stem. This
Figure 4. Five structures of the EIAV TAR RNA are superimposed that contain no violations of the NMR-derived distance and angle constraints, minimum values of the X-PLOR energy function. These are representative of the range of structures that satisfy our NMR-derived constraints equally well. Nucleotides 3-24 are shown. The A-form helical stem is essentially the same in each structure, since the stem was restricted to ideal A-form geometry during the simulated annealing procedure.

forms the same hydrogen bonds as the EIAV TAR RNA U-G base pairs and, like the U10-G17 pair in the equivalent position on the stem, has riboses in C3' endo conformation. Another example is the U-G base pair in the acceptor stem of the yeast tRNA crystal structure (43) which also exhibits the same hydrogen bonding pattern.

Nucleotides C12 and C15. In the structures that are consistent with our NMR data and which have low values for the X-PLOR energy function, the bases of C12 and C15 are stacked upon U11 and G16, respectively (Fig. 5). A comparison of the U11 to C12 and C15 to G16 sequential NOEs with analogous NOEs in the A-form helical stem suggests that their base stacking patterns are similar but not identical. For example, characteristic sequential NOEs are observed from the C12 H6 and H5 base protons to the ribose of U11, as well as an NOE between the U11 H6 and C12 H5 ring protons. The position of C15 is defined by similar sequential NOEs to G16. However, the intensities of these NOEs differ somewhat from the corresponding NOEs in the A-helical stem. Large H1' to H2' scalar couplings indicate that the riboses of C12 and C15 are primarily in the C2' endo conformation.

In our model, the bases of C12 and C15 are oriented towards each other (Fig. 5) and there is the potential for hydrogen bonding to form between them (Fig. 6). Two schemes are likely, and both involve the O2 carbonyl oxygen and nitrogen N3 of C12 as proton acceptors. In the first scheme, the two N4 amino protons of C15 are the donors (Fig. 6B). In the second, one N4 amino proton of C15 hydrogen bonds to N3 of C12 and the N3 of C15 is protonated and bonds to the O2 of C12 (Fig. 6C). C-C pairs have previously been observed in DNA in solution at pH 5.1 (44) and presumably C-C pairing should also be possible in RNA. Unlike the U-G base pairs, we were not able to identify any slowly exchanging protons that are candidates for hydrogen bonds between C12 and C15, at pH as low as 5.0. However, we do not consider this to be significant evidence against the existence of a C-C pairing since this pair is at the end of the helical region and quite accessible to the solvent. It is common, even in Watson-Crick base pairs, for the hydrogen bonded protons at the end of an RNA helix to exchange with the solvent too rapidly for detection by NMR methods. Even if C12 is not protonated, it is also possible to form a C-C pseudo base pair with one hydrogen bond. Base stacking
The bases of U13 and G14 are solvent exposed. This could be accounted for by these protons having fast exchange rates with the solvent, which would be expected if the nucleotides are not particularly stacked on C12 and C15, it is possible that they are stacked one upon the other, and their motion may be stabilized by this base stacking.

**Relation of structure and function**

Carvalho and Derse (28) have performed a detailed structure/function analysis of EIAV TAR by mutating specific nucleotides and assessing the effect on Tat-mediated trans-activation. These experiments were designed and interpreted on the basis of a predicted stem–loop structure for the RNA element, and our initial structural results have shown their conclusions to be largely correct. It is useful to re-examine their results in the light of our improved three-dimensional model of the EIAV TAR. The mutations were of three types, those that disrupt the stem, the U-G base pairs at the top of the stem and the loop conformation. These will be discussed in turn.

Our model clearly shows that the EIAV TAR forms a stable stem–loop structure in which the stem has an A-form helical conformation. Mutations within the stem show that this region cannot be disrupted without losing function but that it can be replaced by another base paired sequence. The role of the stem is clearly to position the distal loop nucleotides such that they can fold into their functional conformation.

Regarding the U-G base pairs (U10-G17 and U11-G16), the conclusion from the mutagenesis experiments was that at least one must be present to maintain some function. In our model, these nucleotides extend the helical stem in a partially distorted A-form helical conformation. There are two possible interpretations of the experiments: either the presence of the base pairs or their identity is important to function. Based on our structure, a simple way of testing this would be to create an EIAV TAR RNA with a perfect A-form helical stem comprising nucleotides 3–11 and 16–24 without the U-G base pairs. Such a molecule ought to have a similar conformation to the wild type structure, unless the nucleotides are not markedly different from the rest of the RNA. This provides some evidence against U13 and G14 being significantly more mobile than the other nucleotides in the RNA loop. While it is clear that the bases of U13 and G14 are not stacked on either C12 or C15, it is possible that they are stacked one upon the other, and their motion may be stabilized by this base stacking.
the C-C pair in the loop are curious and potentially interesting. Simply removing C15 abolishes trans-activation. This deletion would disrupt the C-C pair, which presumably affects the position of the neighboring G14. However, replacing C15 by any other nucleotide maintains or actually increases the wild-type level of trans-activation. Either A, G or U in position 15 can potentially form a base pair with C12 (G-C, A-C mismatch or U-C mismatch), so that replacing C15 with another nucleotide may not substantially alter the presentation of G14. These observations are consistent with a model where C15 does not directly participate in the interactions that lead to the trans-activation effect. Perhaps chemical modification data using DMS which targets the N3 of cytosine could lend further support to a structural model for C12 and C15.

What does the EIAV TAR structure tell us about its relationship to the Tat-TAR system in HIV-1? The biochemistry and structure of the latter has been the subject of intense study and the molecular basis of the interaction has now been rather well characterized. The HIV-1 TAR RNA forms a well-defined stem loop structure with a distinct uridine bulge (12,46,47). The basic region of the HIV-1 Tat protein interacts with the bulge (48) and the loop region is essential for trans-activation (49). It was further shown that for HIV-1 a single arginine can interact with the TAR element and that the binding results in a conformational change within the uridine bulge (47). One uridine enters the major groove to form a base triple and the guanidinium group becomes hydrogen bonded to the guanine of a G-C base pair to form a pseudo triple. It has been suggested that the two U-G base pairs that cap the stem structure in EIAV TAR might be equivalent to the C-C pair in HIV-1 (48) and EIAV (28) and may represent an essential part of the recognition element for an additional cellular factor(s) other than the Tat protein.

REFERENCES